

ŽELJKO BUJAS, ZAGREB

NEKA SVOJSTVA ČESTOTNOG USTROJSTVA
HRVATSKOG LEKSIKA

1.1. U prošlom broju »Filologije«¹ ukazao sam na neke zakonitosti strukture leksika američkog engleskog — prvenstveno na njegovo čestotno ustrojstvo, ali i kulturnoantropološku relevantnost najčešćih imenica i, konačno, na povezanost frekvencijske strukture i etimološkog sastava tog (američko-engleskog) vokabulara. Osim toga, pokazao sam — usporedbom podataka za središnju jezgru američkoengleskog i ruskog vokabulara — da se tu radi o univerzalnim svojstvima evropskih jezika. Nema nikakve sumnje da bi i frekvencijska struktura hrvatskog leksika — da postoji sumjerljivi hrvatski čestotnik — pokazala jednake zakonitosti. Na takav čestotnik morat ćemo počekat i još neko vrijeme, do objavljivanja prvih rezultata rada na Frekvencijskom rječniku hrvatskog književnog jezika (pod rukovodstvom Milana Mogaša, unutar projekta »Korpus hrvatskog književnog jezika« u Zavodu za lingvistiku Filozofskog fakulteta u Zagrebu).

1.2. U međuvremenu, stanoviti uvid u frekvencijsku strukturu hrvatskog leksika pružaju dva mala postojeća čestotnika: Furlanov² iz 1961. i Šojatov³ iz 1983. Oba su izrađena na osnovi korpusa ograničenog opsega — Furlanov sa 100 000, Šojatov s nešto iznad 130 000 riječi teksta.

Korpusi su se razlikovali i po svom sastavu, to jest po prirodi obuhvaćenih tekstova. Furlanov korpus bio je općenitiji, uključivši sljedećih deset područja pismenog izražavanja (svako s po 10 000 riječi teksta):

narodne pripovijetke,
narodne pjesme,
oglasi, reklame, obavijesti,
jedna početnica,
jedan priručnik za lijepo ponašanje,
dnevna štampa,
suvremena prozna književnost,
suvremena poezija,
znanstveni tekstovi
znanstvenopopularni tekstovi

Šojatov korpus imao je za cilj frekvencijsku analizu samo novinskog teksta, pa je sastavljen od ukupno 240 članaka iz »Vjesnika« i »Večernjeg lista« (10 brojeva između 21. ožujka i 1. travnja 1980). Bile su obuhvaćene sljedeće rubrike:

¹ Željko Bujas, »O središnjoj jezgri američkoengleskog vokabulara«, *Filologija* 14 (1986), JAZU, Zagreb, 69—76.

² Ivan Furlan, *Raznolikost rječnika i struktura govora*, neobjavljena disertacija, Sveuč. u Zagrebu, 1961, str. 3—34 Dodatka tezi.

³ Zorislav Šojat, *Čestotni rječnik Večernjeg lista i Vjesnika*, OOUR Agencija za marketing »Vjesnik«, Zagreb, 1983.

vanjska politika	gradska kronika
unutrašnja politika	sport
kultura	crna kronika

1.3. Pokazalo se da Furlanov korpus uključuje 11 399 *različitih* riječi (bez ponavljanja) koje ćemo nazvati »različnicama« (kod Furlana: »riječi razlike«; engleski stručni termin: *types*). Sam korpus sastojao se, kao što znamo, od 100 000 riječi teksta s *ponavljanjima*, koje ćemo nazvati »pojavnica« (kod Furlana: »riječi pojave«; engleski *tokens*). Šojatov korpus imao je 12 569 različenica na svojih 130 279 poavnica.

1.4. Oba projekta prikazuju svoje rezultate u obliku ranglista (Šojatov ima i abecednik te odostražnik). Usporedbom »unutrašnje« čestotne jezgre, to jest 100 jedinica s vrha obje rangliste, dobit ćemo jasan uvid u različitost dva korpusa, ali nesumnjivo i u sličnost (dakle stanovitu zakonitost) leksičke strukture hrvatskog jezika. Konačno, usporedbom tih zajedničkih crta za hrvatski s onim što već znamo o čestotnom ustrojstvu (i njegovoj kulturno-antropološkoj relevantnosti) američkog engleskog i ruskog leksika, doći ćemo do daljnjih spoznaja o ustrojstvu hrvatskog vokabulara, posebno o njegovoj sumjerljivosti s vokabularima suvremenih evropskih jezika.

1.5. Evo dakle paralelnih ranglista prvih sto riječi s Furlanove i Šojatove liste⁴:

FURLAN					ŠOJAT				
(redni Rang broj)	Leksička jedinica	Apsolutna frekvencija (ukupni broj javljanja u korpusu)	Relativna frekvencija (% od ukupnog broja riječi u korpusu)	Kumulativna (pribrojana) relativna frekvencija	Leksička jedinica	Apsolutna frekvencija (ukupni broj javljanja u korpusu)	Relativna frekvencija (% od ukupnog broja riječi u korpusu)	Kumulativna (pribrojana) relativna frekvencija	
1	2	3	4	5	2	3	4	5	
1.	biti	4 483	4,483	—	biti	7 380	5,665	—	
2.	i	3 817	3,817	8,300	i	5 307	4,074	9,379	
3.	u	2 414	2,414	10,714	u	4 832	3,709	13,448	
4.	se	2 356	2,356	13,070	se(be)	2 794	2,145	15,593	
5.	[broj]	1 832	1,832	14,902	da	2 325	1,785	17,378	
6.	da	1 696	1,696	16,598	na	2 052	1,575	18,953	
7.	na	1 462	1,462	18,060	taj	1 843	1,415	20,368	
8.	on	1 400	1,400	19,460	koji	1 706	1,310	21,678	
9.	za	1 031	1,031	20,491	za	1 589	1,220	22,898	
10.	htjeti	892	0,892	21,383	s(a)	1 007	0,773	23,671	

⁴ Obje su liste lematizirane, to jest kosi oblici podvedeni su pod temeljni, kanonski oblik, razriješeni su homografi i slično. Postupak lematizacije proveden je uz manje nedostatke. Furlanu se tako može zamjeriti stvaranje posebne jedinice *ne biti* (29.) kao pretjerani pragmatizam. Kod njega, dalje, susrećemo jedinicu *mi* (35.) nerastavljenu na plural i dativ singulara (podvediv pod *ja*), ili *ti* nerazložen na nominativ jednine lične zamjenice, množinu pokazne zamjenice (= *taj*) i dativ singulara (= *ti*). Kod Šojata, opet, na 75. mjestu nailazimo na oblik *društven* bez diferencijacije na *društven* (= druževan) i *društveni*. A na 100. mjestu navodi se *udružiti*, premda se svi popisani kosi oblici (osim 2) odnose na *udružen* — što je dovoljno samostalna leksička jedinica da bude navedena u tom obliku.

1	2	3	4	5	2	3	4	5
11.	ne	891	0,891	22,274	a	874	0,671	24,342
12.	koji	821	0,821	23,095	o	830	0,637	24,979
13.	a	790	0,790	23,885	od	826	0,634	25,613
14.	s(a)	688	0,688	24,573	htjeti	798	0,613	26,226
15.	od	637	0,637	25,210	sav	717	0,550	26,776
16.	ona	590	0,590	25,800	što	689	0,529	27,305
17.	to	589	0,589	26,389	ne	642	0,493	27,798
18.	što	548	0,548	26,937	godina	622	0,477	28,275
19.	ja	539	0,539	27,476	ne	612	0,470	28,975
20.	kada	509	0,509	27,985	iz	547	0,420	29,395
21.	pa	486	0,486	28,471	ovaj	470	0,361	29,756
22.	svoj	435	0,435	28,906	sam	466	0,358	30,114
23.	moći	428	0,428	29,334	onaj	457	0,351	30,465
24.	taj	397	0,397	29,731	svoj	457	0,351	30,816
25.	jedan	393	0,393	30,124	moći	444	0,341	31,157
26.	kao	375	0,375	30,499	kao	432	0,332	31,489
27.	ti	365	0,365	30,864	on	430	0,330	31,819
28.	svi	361	0,361	31,225	ali	414	0,318	32,137
29.	ne biti	358	0,358	31,583	do	409	0,314	32,451
30.	ili	352	0,352	31,935	jedan	399	0,306	32,751
31.	ali	345	0,345	32,280	rad	390	0,299	33,056
32.	po	323	0,323	32,603	mnogo	362	0,278	33,334
33.	imati	290	0,290	32,893	velik	356	0,273	33,607
34.	tako	287	0,287	33,180	drugi	354	0,272	33,879
35.	mi	276	0,276	33,456	imati	348	0,267	34,146
36.	iz	274	0,274	33,730	trebati	345	0,265	34,411
37.	kako	273	0,273	34,003	naš	326	0,250	34,661
38.	o (<i>prep</i>)	272	0,272	34,275	kako	317	0,243	34,904
39.	drugi	263	0,263	34,538	još	304	0,233	35,137
40.	li	259	0,259	34,797	dobar	287	0,229	35,358
41.	sve	257	0,257	35,054	nov	281	0,216	35,574
42.	velik	249	0,249	35,303	već	273	0,210	35,784
43.	ovaj	246	0,246	35,549	kad	250	0,192	35,976
44.	žena	243	0,243	35,792	sad	248	0,190	36,166
45.	moj	232	0,232	36,031	neki	246	0,189	36,355
46.	Tesla	230	0,230	36,261	ili	244	0,187	36,542
47.	dati	227	0,227	36,488	prvi	244	0,187	36,729
48.	ruka	226	0,226	36,714	Zagreb	237	0,182	36,911
49.	reći	223	0,223	36,937	jer	236	0,181	37,092
50.	onaj	222	0,222	37,159	dva	232	0,178	37,270
51.	do	214	0,214	37,373	organizacija	231	0,177	37,447
52.	ni	214	0,214	37,587	tako	229	0,176	37,623
53.	brat	210	0,210	37,797	dan	227	0,174	37,797
54.	telefon	210	0,210	38,007	pa	213	0,164	37,961
55.	naš	207	0,207	38,214	takav	211	0,162	38,123
56.	kod (<i>prep</i>)	206	0,206	38,420	reći	209	0,160	38,283
57.	samo	205	0,205	38,625	zbog	208	0,160	38,443
58.	već	205	0,205	38,830	ni	202	0,155	38,598
59.	onda	203	0,203	39,033	ostati	200	0,153	38,751
60.	još	202	0,202	39,235	mali	199	0,153	38,904

1	2	3	4	5	2	3	4	5
61.	doći	200	0,200	39,435	po	199	0,153	39,057
62.	rad	194	0,194	39,629	prema	196	0,151	39,208
63.	vi	193	0,193	39,822	zemlja	194	0,150	39,358
64.	svaki	191	0,191	40,013	njegov	191	0,147	39,505
65.	vidjeti	185	0,185	40,198	razvoj	181	0,139	39,644
66.	Zagreb	182	0,182	40,380	dinar	179	0,137	39,781
67.	ono	181	0,181	40,561	oko	179	0,137	39,918
68.	voda	181	0,181	40,742	odnos	178	0,137	40,055
69.	ako	179	0,179	40,921	mi	177	0,136	40,191
70.	jer	176	0,176	41,097	grad	174	0,134	40,325
71.	ići	175	0,175	41,272	predsjednik	167	0,128	40,453
72.	godina	170	0,170	41,442	radnik	167	0,128	40,581
73.	neki	170	0,170	41,612	morati	165	0,127	40,708
74.	nov	169	0,169	41,781	njihov	165	0,127	40,835
75.	prvi	165	0,165	41,946	društven	161	0,124	40,959
76.	dan	162	0,162	42,108	dio	159	0,122	41,081
77.	malen	162	0,162	42,270	dohodak	159	0,122	41,203
78.	dijete	159	0,159	42,429	put	159	0,122	41,325
79.	sada	149	0,149	42,578	vrijeme	156	0,120	41,445
80.	njegov	146	0,146	42,724	nego	155	0,119	41,564
81.	trebati	143	0,143	42,867	prije	155	0,119	41,683
82.	kuća	141	0,141	43,008	mnogi	154	0,118	41,801
83.	te	140	0,140	43,148	pravi	153	0,117	41,978
84.	grad	133	0,133	43,281	radni	151	0,116	42,094
85.	star	131	0,131	43,412	čovjek	150	0,115	42,209
86.	poduzeće	128	0,128	43,540	bez	149	0,114	42,323
87.	znati	128	0,128	43,668	igra	148	0,114	42,437
88.	gdje	126	0,126	43,794	pitanje	139	0,107	42,544
89.	mład	125	0,125	43,919	posto	139	0,107	42,615
90.	glava	124	0,124	44,043	program	138	0,106	42,757
91.	vrijeme	121	0,121	44,164	ako	133	0,102	42,859
92.	bez	120	0,120	44,284	težak	133	0,102	42,961
93.	mnogo	120	0,120	44,404	uz	133	0,102	43,063
94.	dobar	119	0,119	44,523	dalek	132	0,101	43,164
95.	ljudi	118	0,118	44,641	akcija	131	0,101	43,265
96.	više	118	0,118	44,759	film	127	0,098	43,363
97.	stati	117	0,117	44,876	sredstvo	126	0,097	43,460
98.	vrsta	116	0,116	44,992	no	125	0,096	43,556
99.	čovjek	114	0,114	45,106	sud	125	0,096	43,652
100.	otići	111	0,111	45,217	udružiti	125	0,096	43,748

2.1. Već usporedba prvih deset jedinica s obje liste ukazuje na značajne zakonitosti u strukturi hrvatskog leksika, sumjerljive sa i poznate iz sličnih lista za druge evropske jezike:

- (a) Potpuno se poklapa prvih šest jedinica (*biti, i, u, se(be), da, na*). Ova zakonitost već nam je poznata s dvije velike (nelematizirane) rangliste za engleski jezik: tzv. Brown Korpusa⁵ za američki engleski i LOB

⁵ W. Nelson Francis i Henry Kučera, *Computational Analysis of Present-Day American English*, Brown University Press, Providence, R. I., 1967. i (isti autori) *Frequency Analysis of English Usage*, Houghton Mifflin, Boston, Mass., 1982.

Korpusa⁶ za britanski engleski. Budući da su opseg i struktura oba korpusa identični, poklapanje je još značajnije — čak devet od prvih deset jedinica!

- (b) Svih deset najčešćih jedinica čine »gramatičke« a ne »leksičke« riječi. Isto u potpunosti vrijedi za američki i britanski engleski (*the, of, and, to, a, in, that, is, was, he/it*), ali i za ruski⁷ (*v/vo, i, ne, na, ja, byt', čto, on, s/so, a*), njemački⁸ (*der /član/, er, ein, und, in, sein, werden, von, der /zamjenica/, zu*) i francuski⁹ (*de, la, être, et, que, le, a, l', avoir, les*).
- (c) Deset najčešćih riječi u Furlanovoj listi pokrivaju nešto više od petine (21.4%), a kod Šojata gotovo četvrtinu (23.7%) mase korpusa. Razlika stupnja »pokrivenosti« proizlazi iz logične činjenice da veći korpus omogućuje češće javljanje u tekstu najučestalijih riječi. Da to vrijedi i za engleski jasno se vidi iz (lematizirane) rangliste velikog Brown Korpusa američkog engleskog gdje prvih deset riječi pokriva čak 27.3% teksta (v. Ž. Bujas u »Filologiji« 14). Slični su i podaci za već spomenuti korpus njemačkog novinskog teksta (26,958%).

2.2. Dinamizam čestotne strukture hrvatskog leksika bit će nam jasno predočen usporednim kvantitetnim prikazom Furlanove i Šojatove rangliste (u segmentima po deset jedinica):

Rang	Ukupna apsolutna frekvencija		Ukupno pokriivanje Korpusa (u %)	
	Furlan	Šojat	Furlan	Šojat
1— 10	21 383	30 835	21,38	23,67
11— 20	6 602	7 157	6,60	5,49
21— 30	3 950	4 378	3,95	3,36
31— 40	2 862	3 389	2,86	2,60
41— 50	2 362	2 491	2,36	1,91
51— 60	2 076	2 129	2,08	1,63
61— 70	1 862	1 848	1,86	1,42
71— 80	1 627	1 613	1,63	1,24
81— 90	1 319	1 476	1,32	1,13
91—100	1 174	1 266	1,17	0,97

I ovdje je, kao i u analizi američkog engleskog (»Filologija« br. 14), očit drastičan pad kumulativnosti između prve i druge desetine. U daljim desetinama, stopa kumulativnosti kontinuirano, ali sve blaže, opada. Pad akumu-

⁶ Knut Hofland i Stig Johansson, *Word Frequencies in British and American English*, The Norwegian Computing Centre for the Humanities, Bergen, 1982.

⁷ L. N. Zazorina (ed.), *Častotnyj slovar' russkogo jazyka*, Russkij jazyk, Moskva 1977.

⁸ P. Imbes (ed.), *Dictionnaire des Fréquences (Vocabulaire Littéraire des XIXe et XXe Siècles)*, Didier, Paris, 1971.

⁹ I. Rosengren, *Ein Frequenzwörterbuch der deutschen Zeitungssprache*, CWK Gleerup, Lund, 1972 i 1977.

lativnosti posebno je drastičan ako usporedimo prvih i drugih 50 jedinica rangliste (37,16% naprama 8,06% kod Furlana, odnosno 37,27% prema 6,46% kod Šojata).

2.3. Opadanje stope akumulativnosti još je očitije kad se razmotri kvantitativna struktura ranglista po stotinama jedinica. To ćemo ilustrirati samo podacima za Furlanovu listu,¹⁰ za koju postoje podaci o kumulativnoj frekvenciji (već izračunati ili određivi interpolacijom):

TABELA KUMULATIVNOSTI HRVATSKOG LEKSIKA (FURLAN)
ZA PRVIH 1000 JEDINICA (PO 100)

Rang	% pokrivenosti Korpusa	Kumulativna stopa (+ %)
1— 100	45,217	—
— 200	53,238	8,02
— 300	57,924	4,69
— 400	61,377	3,45
— 500	64,055	2,68
— 600	66,290	2,24
— 700	68,240	1,95
— 800	69,971	1,73
— 900	71,513	1,54
—1000	72,902	1,39

Vrijednosti za američki engleski vrlo su bliske (v. »Filologija«, citirani broj, str. 73).

2.4. Konačno, postoje i specifične zakonitosti opadanja kumulativne stope u završnim dijelovima čestotne liste, što je vidno iz Furlanovog čestotnika razmotrenog po pojedinim tisućama jedinica:

TABELA KUMULATIVNOSTI HRVATSKIH LEKSIKA (FURLAN)
PO 1000 JEDINICA

Rang	% pokrivenosti Korpusa	Kumulativna stopa (+ %)
1—1000	72,902	—
—2000	82,085	9,18
—3000	88,082	6,00
—4000	90,329	2,25
—5000	92,540	2,21
—6000	94,540	2,00
—7000	95,540	1,00
—8000	96,540	1,00
		(dalje isto)

Kako vidimo, nakon sedme tisuće kumulativna stopa postaje konstantna, odražavajući poznato svojstvo vokabularskih ranglista — da naime donju

¹⁰ Šojatova ranglista ne sadrži nažalost te podatke, pa sam one za prvih 100 jedinica ručno izračunavao (što je prilično mukotrpan postupak).

polovicu liste čine riječi frekvencije 1. To je posebno izraženo kod velikih korpusa. Tako ukupni broj riječi s frekvencijom 1 u Brown Korpusu iznosi 22 543 (od ukupno 50 406 različenica), a pokrivaju 44,7% korpusnog teksta. (O značajnim implikacijama opadanja kumulativne stope nakon treće tisuće za učenje stranih jezika vidi u »Filologiji« br. 14, str. 74 i 75.)

3.1. Frekvencijska struktura vokabulara sadrži međutim i druga još značajnija, svojstva i zakonitosti. Vidjeli smo da se čestotna jezgra hrvatskog vokabulara (100 najučestalijih različenica) sastoji pretežno od »gramatičkih« riječi. »Leksičkih« jedinica (»punih riječi«) ima osjetno manje.

Tako Furlanova ranglista (v. tabelu pod 1.3.) sadrži ove leksičke jedinice među prvih sto riječi:

IMENICE *žena* (na 44. mjestu), *Tesla* (46.), *ruka* (48.), *brat* (53.), *telefon* (54.), *rad* (62.), *Zagreb* (66.), *voda* (68.), *godina* (72.), *dan* (76.), *dijete* (78.), *kuća* (82.), *grad* (84.), *poduzeće* (86.), *glava* (90.), *vrijeme* (91.), *ljudi* (95.), *vrsta* (98.) i *čovjek* (99.)

GLAGOLI *biti** (1.), *htjeti** (10.), *moći* (23.), *imati* (33.), *dati* (47.), *reći* (49.), *doći* (61.), *vidjeti* (65.), *ići* (71.), *trebati* (81.), *znati* (87.), *stati* (97.) i *otići* (100.) [Zvjezdicama su označeni glagoli pretežno »gramatičke« funkcije]

PRIDJEVI (OPISNI)

velik (42.), *nov* (47.), *malen* (77.), *star* (85.), *mlad* (89.) i *dobar* (94.)

Šojatova tabela, nužno odražavajući jezik i tematiku dnevnih novina, sadrži sljedeće »negramatičke« riječi među 100 najučestalijih:

IMENICE *godina* (18.), *rad* (31.), *Zagreb* (48.), *organizacija* (51.), *dan* (53.), *zemlja* (63.), *razvoj* (65.), *dinar* (66.), *oko* (67.), *odnos* (68.), *grad* (70.), *predsjednik* (71.), *radnik* (72.), *dohodak* (77.), *put* (78.), *vrijeme* (79.), *čovjek* (85.), *igra* (87.), *pitanje* (88.), *program* (90.), *akcija* (95.), *film*, (96.), *sredstvo* (97.) i *sud* (99.)

GLAGOLI *biti** (1.), *htjeti** (14.), *moći* (25.), *imati* (35.), *trebati* (36.), *reći* (56.), *ostati* (59.) i *morati* (73.)

PRIDJEVI (OPISNI)

velik (33.), *dobar* (40.), *nov* (41.), *mali* (60.), *društveni* (75.), *pravi* (83.), *radni* (84.), *težak* (92.), *dalek* (94.), *udružen* (100.)

3.2. Najučestalije »pune« riječi iz dva razmatrana hrvatska čestotnika nesumnjivo odražavaju različitost korpusa na kojima su ti čestotnici zasnovani: Furlanov koji, unatoč skromnom opsegu, nastoji obuhvatiti najširi raspon sadržaja i stilova i Šojatov, stilski uži jer se ograničava na jezik i tematski raspon dnevnog tiska. Stanoviti broj leksičkih jedinica zajednički je međutim i jednom i drugom korpusu. Za te jedinice mogli bismo dakle uvjetno reći da su po svojoj prirodi univerzalne, prisutne u svakom generalnom, tematski i stilski »prosječnom« hrvatskom korpusu. To su:

imenice — *rad, godina, dan, grad, vrijeme i čovjek*
glagoli — *biti*, htjeti*, moći, imati, trebati i reći*
pridjevi — *velik, nov i malen/mali*

3.2.1. Veći, prema tome i pouzdaniji, korpusi hrvatskih tekstova nesumnjivo bi smanjili broj »punih« riječi unutar prvih sto jedinica čestotne liste na jednu trećinu. Veliki korpusi omogućuju naime »gramatičkim« riječima da se ponavljaju razmjerno opsegu korpusa, penjući se sve više na ranglisti i potiskujući »pune« riječi na niže čestotne položaje. Istovremeno, među »leksičkim« jedinicama izbijaju na sam vrh one koje — posebno imenice — izražavaju koncepte što su u temelju čovjekovih preokupacija: Čovjek sam i njegovo mjesto u vremenu i prostoru. Tu kulturnoantropološku relevantnost frekvencijskog ustrojstva vokabulara dosljedno pokazuju veliki, a sumjerljivi, korpusi američkog i britanskog engleskog, ali i ruskog i drugih evropskih jezika (v. »Filologija« 14, str. 75). Na njihove vrhove naime redovito izbijaju ove imenice:

ENGLESKI *man, people, time, year, day, way, world, state*

RUSKI *god, delo, vremena, čelovek, ljudi, ruka, žizn', den', tovarišč, rabota, glaz*

FRANCUSKI *homme, jour, chose, temps, vie, femme*

3.3. Ništa manje zanimljive nisu imenice po kojima se Furlanov i Šojatov popis razlikuju. To su u Furlanu:

IMENICE *žena, Tesla, ruka, brat, telefon, voda, dijete, kuća, poduzeće, glava, ljudi, vrsta*

GLAGOLI *dati, doći, vidjeti, ići, znati, stati, otići*

PRIDJEVI *star, mlad*

A u Šojatu:

IMENICE *organizacija, zemlja, razvoj, dinar, oko, odnos, predsjednik, radnik, dohodak, put, igra, pitanje, program, akcija, film, sredstvo, sud*

GLAGOLI *ostati, morati*

PRIDJEVI *društveni, pravi, radni, težak, dalek, udružen*

Prisutnost jedinica *Tesla* i *telefon* među 100 najčešćih riječi na Furlanovoj listi nesumnjivo ukazuju na nedovoljnu disperziju uzoraka, što je kao propust omogućilo ovim usko tematskim riječima da izbiju tako neprirodno visoko na ranglisti. (*Tesla* se naime javlja čak 230, a *telefon*, 210 puta!)

¹¹ Željko Bujas, »Cultural-Anthropological Relevance of Word Frequency Parameters (American English vs. British English)«, *SRAZ*, Zagreb, 1988 (u tisku).

Odmah se zapaža i tematska uvjetovanost imeničke liste iz dnevnog tiska (Šojat) — *organizacija, zemlja, dinar, predsjednik, radnik, dohodak* — s time da su neke od njih tipične političke krilatice: *razvoj, program, akcija* (v. tako i za *buzz words* u američkom engleskom /ž. Bujas¹¹, u SRAZ-u za 1988/). Slično je i s pridjevima iz Šojatove liste: *društveni, radni, udružen*. Glagoli su, na prvi pogled, najbezličniji ali visoko mjesto *trebati* na toj listi — odmah nakon frekvencijski vrlo podudarnih *biti* (1./1.), *htjeti* (10./14.), *moći* (23./25.) i *imati* (33./35.) — jasno svjedoči o volontarističkom i forumskom tonu jugoslavenskog političkog diskursa.

3.4. Vratimo se, na kraju ovih poredbenih analiza, i našim bezličnim »gramatičkim« riječima. Začudit ćemo se kako one mogu biti itekako ekspresivne. Ništa ne odražava impersonalniju prirodu novinskog teksta bolje od znatno, drastično nižeg čestotnog ranga ličnih zamjenica u tom tekstu.

	Rang kod Furlana	Rang kod Šojata
<i>ja</i>	19.	121.
<i>ti</i>	27.	202.
<i>on</i>	8.	27.
<i>ona</i>	16.	194.
<i>ono</i>	67.	— (!)
<i>mi</i>	35.	65.
<i>vi</i>	63.	192.
<i>oni</i>	—	—

(Blaža razlika u rangu *mi* proizlazi iz njegove česte funkcije bezličnog izražavanja u svim tipovima teksta.)

Naravno, usporedbe *unutar* korpusa mogu biti jednako indikativne. Tako, više mjesto muške zamjenice *on* od ženske *ona* u oba korpusa (Furlan 8 : 16, Šojat 27 : 194) nepobitno svjedoči o dominaciji muškog u jeziku.

4.1. Analiza daljeg (ne)podudaranja Furlanove i Šojatove liste, na nižim čestotnim pozicijama (u drugoj, trećoj stotini i tako dalje), pružila bi dodatne podatke o općoj strukturi hrvatskog leksika, o specifičnim leksicima kao što je novinarski, o frekvenciji kao parametru stilske vrijednosti leksičke jedinice, i slično. Sve te, i više drugih predvidivih analiza imat će međutim pouzdanije rezultate, omogućit će sigurnije iskaze o čestotnom ustrojstvu i zakonitostima hrvatskog leksika, ako im kao osnova posluži velik i brižljivo sastavljen, doista reprezentativan korpus suvremenog hrvatskog jezika. Pričekajmo dakle dok najavljeni Mogušev čestotnik (kompjuterski pripreman, što je dodatna vrijednost) ne bude pružen na uporabu hrvatskoj stručnoj, znanstvenoj i kulturnoj javnosti.

Summary

SOME PROPERTIES OF THE FREQUENCY STRUCTURE OF CROATIAN LEXIS

Using two existing rank lists of Croatian, based on limited-size corpora of general texts (by I. Furlan) and of newspaper articles (by Z. Šojat), this paper offers a comparative analysis of the core vocabularies (100 top-frequency items) provided by the two lists. Several laws governing the lexical structure of Croatian are determined and discussed, with parallels to English and Russian lexis pointed out. Attention is also drawn to the cultural-anthropological and stylistic relevance of some properties of the frequency structure of Croatian core vocabulary.